

Signatures of Ecological Resource Availability in the Animal and Plant Proteomes

James J. Elser,* William F. Fagan,† Sankar Subramanian,*‡ and Sudhir Kumar*‡

*School of Life Sciences, Arizona State University; †Department of Biology, University of Maryland; and

‡Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University

Although substantial and ecologically significant differences in elemental composition are well documented for whole organisms, little is known about whether such differences extend to lower levels of biological organization, such as the elemental composition of major molecules. In a proteome-scale investigation of 9 plant genomes and 9 animal genomes, we find that the nitrogen (N) content of plant proteins is lower than that in animal proteins. Furthermore, protein N content declines with the intensity of gene expression for plants, whereas the N content of animal proteins shows no consistent pattern with expression. Additional analyses indicate that the differences in N content between plant and animal proteomes and in plant proteins as a function of gene expression cannot be attributed to protein size, GC content, gene function, or amino acid properties. These patterns suggest that ecophysiological selection has operated to conserve N in plants via decreased reliance on N-rich amino acids. This inference was supported by an analysis of conserved and variable sites indicating that the N content of plant amino acids coded by variable sites is similar to that of the sites conserved between plant and animal genomes and shows no association with expression level. In contrast, in animals, the N content of amino acids coded by variable sites is significantly higher than that for conserved sites, suggesting relaxation of selective constraints for N usage in the animal lineage. This constitutes the first evidence for an influence of environmental resource availability on proteomes of multicellular organisms.

Introduction

Unravelling the connections between genomic structures and the ecological interactions among organisms and their environments is a fundamental axis of integration in modern biology. Recent biochemical investigations of microorganisms have revealed a surprising impact of ecophysiological constraints on protein sequences in the form of significant biases in amino acid use according to energy and nutrient element costs (Baudouin-Cornu et al. 2001; Akashi and Gojobori 2002). For example, in *Saccharomyces cerevisiae*, proteins involved in nitrogen (N) transport and metabolism are disproportionately enriched in amino acids that contain few N atoms (Baudouin-Cornu et al. 2001). (Although all 20 amino acids contain at least 1 N atom in their amino group, 6 of them [asparagine, glutamine, lysine, tryptophan, histidine, and arginine] contain an additional 1–3 N atoms in the side chain.) These features are thought to minimize the organism's demand for certain elements at times when these nutrients are most limiting to growth and reproduction (Baudouin-Cornu et al. 2001).

Whether such resource limitations have shaped the proteomes of multicellular organisms remains unknown. It seems possible that the composition of animal and plant proteomes has been shaped by nutrient constraints because such species commonly experience deficiencies and overabundance of key nutrient elements in nature (White 1993; Aerts and Chapin 2000; Elser et al. 2000; Sterner and Elser 2002). In fact, relative to plants, animal biomass features substantially higher nitrogen (N) content (Elser et al. 2000), reflecting major differences in how these groups acquire resources and then allocate overall biomass to low- versus high-N biomolecules (e.g., carbohydrates vs. proteins). But, might differences in N use between plants and animals also be seen in the elemental composition of

the proteins themselves? To evaluate this ecologically motivated hypothesis, we examined the amino acid compositions of all known proteins encoded by completely sequenced genomes of 9 animals and 2 plants, along with data for 7 other plant species for which extensive gene sequence information was available. Our results indicate that indeed the proteomes of the plant taxa are composed of amino acids with significantly lower N usage than the animal proteomes. Furthermore, we find that protein N content is a function of gene expression intensity in plants but not in animals. We suggest that the nature of this functional relationship may differ among taxa due to differences in how the amino acids needed to build proteins are acquired.

Materials and Methods

Estimation of Proteome Elemental Contents

Protein sequences for the species examined were obtained either from Ensembl (<http://www.ensembl.org>) and Unigene (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) data banks. The proteomes were selected based on the availability of their corresponding gene expression data. We also selected the animal proteomes in such a way so as to minimize their phylogenetic dependence with respect to their amino acid contents. For instance, including many mammalian species to represent animal genomes might introduce a bias in the estimation of proteomic nitrogen content due to shared ancestry. The nitrogen (N) content of each proteome, specifically the number of N atoms per amino acid side chain, was estimated as:

$$N_C = (100/L) \sum w_i \times p_i,$$

where w_i is the number of nitrogen atoms in the amino acid side group i (Range = 0–3), p_i is the frequency of the i -th amino acid, and L is the sequence length. We used the Mann–Whitney U test (Sokal and Rohlf 1994) to determine the statistical significance of the difference between plant and animal N_C scores. Likewise, the equation above can be applied to calculate the number of N atoms per

Key words: proteome, nitrogen, animal, plant, stoichiometry.

E-mail: s.kumar@asu.edu.

Mol. Biol. Evol. 23(10):1946–1951. 2006

doi:10.1093/molbev/msl068

Advance Access publication July 26, 2006

nucleotide by assigning weights of 5, 5, 3, and 2 for adenine, guanine, cytosine, and thymine, respectively.

Estimation of Gene Expression Intensity

For each protein, relative gene expression intensity was determined using the expressed sequence tag (EST) counts following previously described BlastN procedures (Duret and Mouchiroud 1999; Subramanian and Kumar 2004). The EST data were obtained either from the dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>) or Unigene databases, and the ESTs from all tissue libraries were pooled to avoid any bias introduced by the expression of tissue-specific genes. Because EST library sizes and the number of genes to which ESTs can be mapped varied among species, we standardized the EST count (E_i) for each expression intensity category, i , as

$$E_i = (100/E) \left(\sum_{1-i} \sum_{1-G_i} e_{ij} \right),$$

where E is the total number of mapped ESTs, G_i is the number of genes in category i , and e_{ij} is the number of ESTs mapped to gene j in the i th category. Using the number of ESTs per gene as a measure of gene expression intensity produced results similar to those reported in figures 1 and 2, except that the scales used in the graphs varied among taxa due to differences in the number of genes in the genomes and the number of ESTs in the libraries used. Therefore, relative gene expression intensity (rescaled to a common axis) was preferred for visual representation and was used in regressions against protein N content or carbon:nitrogen ratio in 2 fully sequenced plant taxa (*Arabidopsis thaliana* and *Oryza sativa*) and 2 selected animal taxa (*Drosophila melanogaster* and *Anopheles gambiae*). However, examination of protein N content as a continuous function of gene expression was not possible for the remaining plant taxa due to the relatively small number of genes available for analysis. To overcome this, we sorted all genes for each of the 18 taxa, both plant and animal, into 2 categories of expression: high (top 3% of EST counts) and low (1 EST). We then compared protein N content in the high and low expression classes for all 18 species considered.

Analysis of Orthologous Sequences

To analyze the elemental composition of orthologous sequence sets between species pairs, putative orthologous relationships were identified using a local BlastP search with BLOSUM62 substitution matrix (Altschul et al. 1997) following a previously outlined reciprocal BlastP search procedure (Subramanian and Kumar 2004). In this approach, the lineage-specific duplicate coorthologous sets (Sonnhammer and Koonin 2002) from both species are accounted for appropriately to include only 1 coorthologous sequence pair in the final data set, and a pair of genes is considered orthologous only if each gene is mutually the best match in its respective counterpart genome (Waterston et al. 2002). *Arabidopsis thaliana* protein sequences and their putative orthologs from *O. sativa*, *D. melanogaster*, and *A. gambiae* were aligned using ClustalW with default settings (Thompson et al. 1994).

Table 1
Nitrogen (N) Content and C:N Ratios in Animal and Plant Proteomes

Species	Number of Genes	N Atoms per Side Chain (SE)	C:N
Plants			
<i>Arabidopsis thaliana</i>	25,544	0.364 (0.000)	8.1
<i>Chlamydomonas reinhardtii</i>	429	0.348 (0.008)	8.1
<i>Glycine max</i>	721	0.353 (0.006)	8.3
<i>Hordeum vulgare</i>	733	0.349 (0.006)	8.1
<i>Lycopersicon esculentum</i>	830	0.345 (0.006)	8.5
<i>Oryza sativa</i>	24,967	0.367 (0.002)	7.9
<i>Solanum tuberosum</i>	516	0.339 (0.006)	8.7
<i>Triticum aestivum</i>	889	0.356 (0.006)	7.9
<i>Zea mays</i>	1,205	0.359 (0.006)	7.9
Overall mean (SE)		0.353 (0.003)	8.2
Animals			
<i>Anopheles gambiae</i>	14,366	0.386 (0.002)	7.7
<i>Caenorhabditis elegans</i>	18,253	0.366 (0.002)	8.3
<i>Drosophila melanogaster</i>	12,516	0.387 (0.002)	7.6
<i>Danio rerio</i>	30,782	0.377 (0.002)	7.9
<i>Gallus gallus</i>	28,415	0.376 (0.002)	7.9
<i>Homo sapiens</i>	34,090	0.383 (0.002)	7.8
<i>Mus musculus</i>	32,280	0.381 (0.002)	7.8
<i>Bos taurus</i>	4,201	0.371 (0.001)	8.0
<i>Xenopus laevis</i>	10,019	0.373 (0.002)	8.0
Overall mean (SE)		0.378 (0.002)	7.9
%Difference			
(Animals – Plants)/Plants		7.08%	–3.70%
P Value			
(Mann–Whitney U test)		<0.0001	<0.02

NOTE.—C:N is the carbon:nitrogen; and SE, standard error.

Results and Discussion

We found that the animals used $\sim 7\%$ more N atoms per amino acid residue in their proteomes than the plants studied (table 1). For example, *D. melanogaster* shows 6.3% more N atoms per side chain than *A. thaliana* (0.387 vs. 0.364), whereas it shows almost identical N content with that seen in *Homo sapiens* and *A. gambiae* (0.383 and 0.386, respectively). Of the 9 animals considered, only *Caenorhabditis elegans* has a proteomic N content similar to that of the plant taxa. In fact, excluding *C. elegans*, the distributions of proteomic N content of the plants and animal taxa are entirely nonoverlapping (0.367 N atoms per side chain in the highest plant taxon vs. 0.371 for the lowest animal proteome; table 1).

Even though the differences observed between the animal and plant taxa are small in absolute magnitude, they are highly significant ($P < 10^{-4}$ in Mann–Whitney U test; table 1 N_p values). This difference is not attributable to differences in the sizes of gene families in these 2 eukaryotic kingdoms or to the presence of lineage-specific proteins because an analysis restricted to putatively orthologous proteins from the complete proteomes of *A. thaliana*, *O. sativa*, *D. melanogaster*, and *A. gambiae* also demonstrated a significant difference ($P < 10^{-9}$; see Supplementary Table 1, Supplementary Material online).

It is also possible that the observed difference reflects genomic GC bias because there is a correlation between the

GC content of the first 2 codon positions and the N content of their associated amino acids in the standard genetic code table (Baudouin-Cornu et al. 2004; Bragg and Hyder 2004). To assess this possibility, we analyzed the GC content of introns in *A. thaliana* and *D. melanogaster*. The GC content of *A. thaliana* introns is 32%, whereas that of *D. melanogaster* introns is 37%; these both differ substantially from the GC content of mouse (45%), but the estimated proteomic N content of fruit fly and mouse proteomes are nearly identical (0.387 and 0.381, respectively). Therefore, it is unlikely that the differences in GC content are responsible for the observed differences among taxa.

Alternatively, the differential N usage between animal and plant proteomes might be merely an outcome of species' differential use of hydrophilic amino acids because 5 of the 6 amino acids that contain additional N atom(s) in their side chains are hydrophilic (histidine, arginine, asparagine, glutamine, and lysine). First, if hydrophilicity is the cause of the observed pattern, then the proportion of hydrophilic amino acids that contain no additional N atoms in their side chains (aspartic acid, glutamic acid, serine, and threonine) should also be low in plant proteomes and high in animal proteomes. We found the opposite pattern: *A. thaliana* used more of these 4 non-N-containing hydrophilic amino acids than did *D. melanogaster* (26.0% vs. 24.8%, respectively). Second, it is known that smaller proteins have a higher surface area to volume ratio and that surface sites are enriched with hydrophilic residues (Akashi and Gojobori 2002). If protein size was driving the signal in proteomic N use, this would lead to higher N content and smaller proteins in animals (as most of the residues with a N-containing side chain are hydrophilic). In contrast, the plant proteins were slightly, but significantly ($P < 10^{-8}$), smaller than the animal proteins on average (495 vs. 520 amino acids, respectively).

One might also propose that the differences in N content between plant and animal proteomes reflect differences in the dominance of proteins of distinctly different functions, as would be true if N content is correlated with distinct functional properties of particular amino acids used in proteins unique to animal or plant proteomes. To examine this, we conducted an analysis for *A. thaliana* and *D. melanogaster* that included only proteins belonging to the same functional group (using KOG database, <http://www.ncbi.nlm.nih.gov/COG/new/>) and still found a significant difference in proteomic nitrogen content ($P < 10^{-19}$; see Supplementary Table 2, Supplementary Material online).

Alternatively, if the observed difference between animal and plant proteomic N contents is caused by natural selection operating on the efficiency of N use (as was argued for microbial proteomes, Baudouin-Cornu et al. 2001), then we would expect that N-content differences between plant and animal taxa would be greatest for the most highly expressed proteins, which, logically, are likely to be under the strongest selection pressures for N conservation. To test this idea, we examined the relationship of the N content with the relative expression intensity of the corresponding genes in EST libraries. We focused on relative, rather than absolute, values of EST counts because the absolute estimates of gene expression intensities are not precise and vary among libraries and among species. On the other

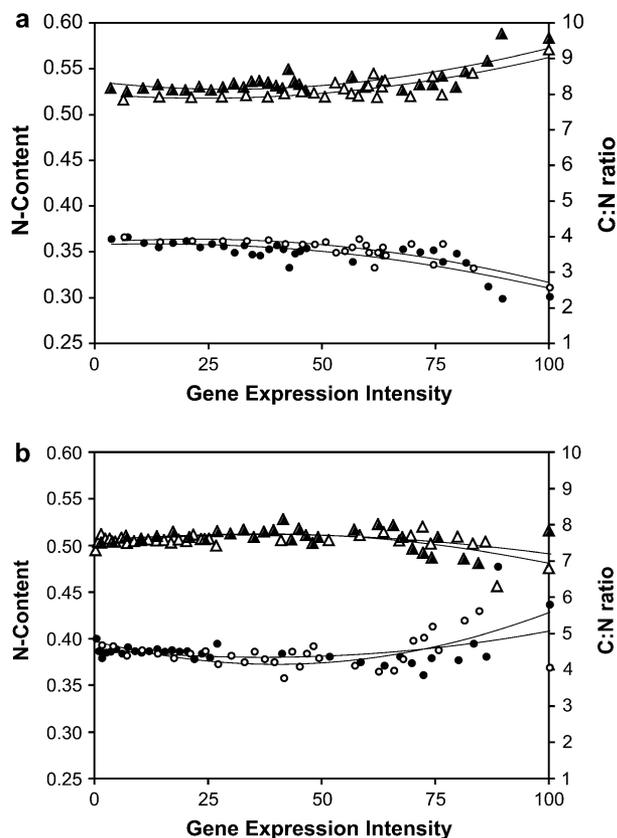


FIG. 1.—N content (circles) and carbon:nitrogen ratio (triangles) of proteins as a function of relative expression intensity of the genes that encode them. Data are shown for (a) 2 plant species (mustard weed, *Arabidopsis thaliana*, dark; and rice, *Oryza sativa*, open) and (b) 2 animal species (fruit fly, *Drosophila melanogaster*, dark; and mosquito, *Anopheles gambiae*, open). Solid lines show the fit of a second-degree polynomial to each pattern. In all analyses, the explanatory power (R^2) of the quadratic regression exceeded that of the related linear regression, and the quadratic fit was statistically preferred over the related linear regression even after accounting for the increased parametric flexibility of the quadratic fits via Mallows Cp (Hilborn and Mangel 1997). The quadratic regressions for mosquito were close, but nonsignificant.

hand, it is well known that different techniques for measuring relative gene expression levels (such as ESTs, and microarrays) yield similar results and there is high correlation from diverse species (e.g., mouse and *Drosophila*) in expression levels of the same genes (Lercher et al. 2002; Subramanian and Kumar 2004). Alternatively, codon usage bias of the genes could be used as a proxy for the gene expression level as they are known to be highly correlated (Akashi 2003). However, existence of such a relationship has not been confirmed in vertebrates due to their smaller population sizes (Akashi 2003).

We found that N content decreased with increasing expression in both *A. thaliana* and *O. sativa* (fig. 1a): proteins coded by the most highly expressed genes showed a $\sim 15\%$ lower N content than the average over all genes ($P < 10^{-3}$). In contrast, *D. melanogaster* and *A. gambiae* proteins encoded by very highly expressed genes appeared to have a somewhat higher N content (fig. 1b). When we extended the analysis to the other species in our study, we found that increased protein N content with expression is not a general

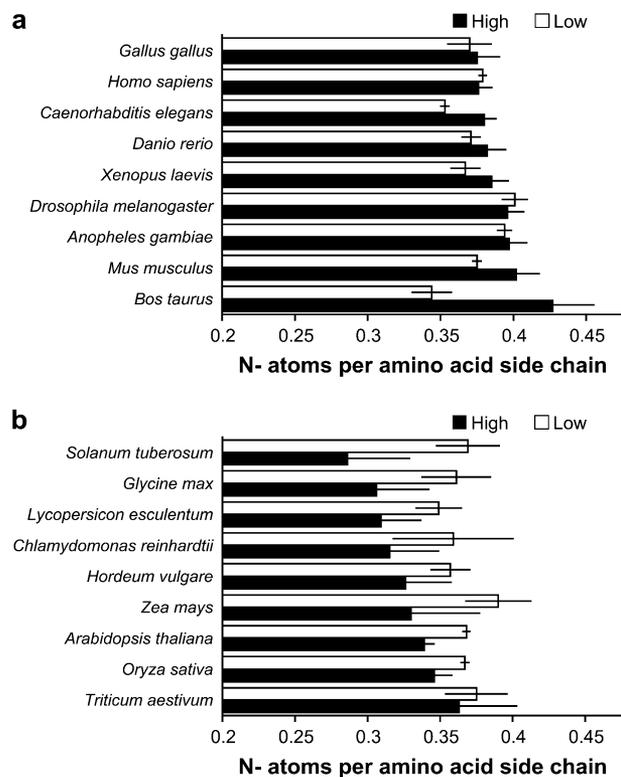


FIG. 2.—Nitrogen content of amino acid side chains of proteins from high and low expression genes for 9 animal (a) and 9 plant (b) taxa. Genes were sorted based on their expression level, and the top 3% were considered high expression genes. Genes having only one EST were classified as low expression. The error bars indicate the standard error of the mean.

property of animal proteomes (fig. 2a). Whereas the N content of some of the animal proteomes showed a significant increase with expression level, for other animal taxa protein N contents were not significantly different between the expression level categories. Further analysis of the animal proteomes revealed that the elevated N content of highly expressed genes was primarily due to the presence of particularly N-rich ribosomal proteins that constitute 10–20% of the proteins in this category. When this specific group of proteins was excluded, the difference between the N content of high and low expressed proteins of these animal genomes disappeared. However, high expression proteins had lower N content than low expression proteins in all of the 9 plant taxa (fig. 2b), suggesting that this pattern is a general one for plants. Thus, relative to the overall proteomes, the difference in animal and plant proteomic N content was considerably larger for highly expressed proteins (19%; 0.391 ± 0.005 standard error vs. 0.329 ± 0.007 , respectively). It should be noted that the overall proteomic N contents of completely sequenced plant proteomes (0.364–0.367) were slightly higher than that of the plants with more limited protein data sets (0.339–0.359) (table 1; additional information is provided in Supplementary Table 3, Supplementary Material online). This difference likely results from the fact that, due to preferential sequencing, the partial genomes contain a relatively greater proportion of highly expressed genes than the complete

genomes, resulting in a somewhat lower estimated proteomic N content.

Our results suggest that the ecophysiological observation of lower N content in plant biomass relative to animal biomass (Elser et al. 2000) also applies to entire proteomes, although the proteomic difference per amino acid is much smaller than the difference in the overall biomass. In addition, natural selection appears to be playing a significant role in shaping proteomic amino acid composition, as evidenced by the interaction between relative gene expression intensity and protein N usage.

Before discussing the ecophysiological interpretation of these patterns, we first reevaluate the possibility that the GC content and the differential use of hydrophilic amino acids is causing the observed pattern. We do this because we have shown above that a comparison involving genes with higher expression intensities provides much more power in rejecting the null hypothesis (i.e., a 19% difference in plant and animal N usage in high expression proteins vs. 7% for all proteins). We evaluated whether the GC content of genes covaried with gene expression intensity in plant genomes because of the inherent relationship between the GC content of codons and the N content of the amino acids coded by them (Baudouin-Cornu et al. 2004; Bragg and Hyder 2004). If such a relationship is present in plant but absent in animal genomes, then the observed difference in N content of high and low expression plant proteins might be explained by variation in genomic base composition alone. Our analyses reveal that, in the case of *A. thaliana*, the proteomic N content was slightly higher for genes in GC-rich relative to GC-poor regions (0.346 and 0.356 for proteins from genes with intronic GC contents <30% and >35%, respectively). However, a similar pattern was also observed for *D. melanogaster* (0.374 and 0.385 N atoms per side chain for intronic GC contents <30% and >40%, respectively). Because N content varied in the same direction with GC content in both plant and animal proteomes, our result of strong trends in N content with expression in plants but not animals cannot be explained by a bias in base composition (see Supplementary Table 4, Supplementary Material online for more details). Furthermore, in *A. thaliana* the GC content in high expression genes did not differ significantly ($P = 0.1$) from GC content in low expression genes. Therefore, the variation in protein N content as functions of taxon and of expression intensity cannot be attributed to trends in nucleotide base composition.

As mentioned earlier, of the 6 N-containing amino acids all but tryptophan are hydrophilic. Therefore, if proteins from highly expressed plant genes contain fewer hydrophilic amino acids than their respective proteins from weakly expressed genes, then hydrophilicity alone could explain the observed pattern in plants. This possibility can be examined by estimating the proportion of non-N-containing hydrophilic amino acids (aspartic acid, glutamic acid, serine, and threonine). Under the alternative hypothesis just described, we expect a reduction of these amino acids in proteins from highly expressed genes of plants and no such relationship in the case of animals. However, this alternative can also be rejected because the proportion of these other hydrophilic amino acids was a decreasing function of gene expression intensity not only in *A. thaliana*

($P < 10^{-10}$) but also in *D. melanogaster* ($P < 10^{-4}$) proteomes, suggesting a different kind of causal mechanism that is common to both taxa. Similar results were obtained when we analyzed only the aromatic amino acids (2 of the 6 N-containing amino acids have aromatic side chains). Significant differences in the aromatic amino acid content of high and low expression proteins were observed for *A. thaliana* ($P < 10^{-6}$) as well as *D. melanogaster* genomes ($P < 10^{-7}$). These analyses suggest that biases in uses of these kinds of amino acids (hydrophilic and aromatic) in proteins associated with high and low expression genes are similar across the 2 kingdoms. Thus, the difference in proteomic N content observed for plants and animals is independent of the amino acid properties, suggesting a possible role for selection for overall efficiency of N usage in driving the pattern.

Further support for this “N efficiency” interpretation derives from analyses of conserved and variable sites in the orthologous proteins of *A. thaliana*, *O. sativa*, *A. gambiae*, and *D. melanogaster*. First, the N content of the conserved protein sites (sites that are identical in plant and animal genomes) is not significantly different from the variable sites ($P > 0.05$) of the plant proteomes, suggesting that plants have maintained a lower protein N content since the time of the last shared ancestor of plants and animals. However, in animals, the N content of the variable amino acid positions is significantly higher than that of the conserved position ($P < 10^{-5}$). These results are a clear indication that protein evolution in plants has undergone selection for efficient protein N use, whereas in animals, such selection pressures have been either relaxed or counteracted by other selective forces.

We suggest 2 evolutionary mechanisms for the observed patterns. First, the lower N content of plant proteins and its decline with expression levels likely reflect the fact that plants, like most microbes, retain primary amino acid synthesis pathways in which the entire suite of amino acids is synthesized from raw materials. Therefore, when environmental N supplies are limiting, reduced reliance on high-N amino acids allows N to be allocated to other uses in the cell, benefiting overall growth and reproduction (Stern and Elser 2002). Second, because animals do not directly use inorganic N from the environment and instead obtain preformed amino acids from their diets (Lehninger et al. 1993), purifying selection acting on efficiency of N use is reduced. Indeed, the incorporation of N-rich amino acids obtained from the diet into the proteome may be favored because of enhanced translational efficiency (Akashi 2003), a mechanism well known to influence codon frequencies.

In contrast to the above-mentioned mechanisms relating to the under- or overabundance of amino acid residues available to build proteins, it is conceivable that selection has operated not via mechanisms associated with N conservation in proteins but instead via possible effects of N limitation on codon use during transcription. This is possible because the N contents of codons and their associated amino acids are theoretically correlated based on the genetic code (Bragg and Hyder 2004). However, this can be rejected because the genome-wide average N content of the actual codons used was nearly identical for genes

expressed with the lowest and the highest intensities in *A. thaliana* (3.780 and 3.777 N atoms per nucleotide, respectively, $P = 0.57$; 11.30 and 11.34 per codon, respectively, $P = 0.47$) and in *D. melanogaster* (3.825 and 3.836 N atoms per nucleotide, respectively, $P = 0.12$; 11.51 and 11.50 per codon, respectively, $P = 0.49$). It is also possible that the biases we report are driven by complex relations related to the biochemical properties of proteins that differentially affect low- versus high-N amino acids. This possibility is not supported by our comparison of plant and animal proteins sorted into similar functions, but additional studies of protein N use, amino acid investment, and architecture are needed, especially as more plant genomes become available for analysis.

Our findings provide the first suggestion that the eco-physiological footprints of resource limitations can be seen not only in microbial proteomes (Baudouin-Cornu et al. 2001) but also in those of higher organisms. They also indicate that the evolutionary “fitness” of various amino acids may differ depending on whether those amino acids are found in an animal or a plant and whether they are associated with a high expression or a low expression gene. Although the evolutionary underpinnings of these patterns require more investigation, our results suggest that even small overall differences in amino acid composition of proteomes may be linked to environmental constraints on the organism because small differences may become greatly magnified when viewed in the context of the intensity of gene expression. This provides motivation for similar hypothesis-driven investigations that consider species showing much larger differences in proteomic chemical composition as these are likely to provide further evidence for links between a species’ proteome and its environment and ecology.

Supplementary Material

Supplementary Tables 1–4 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank S. Gadagkar, J.P. Collins, C. Delwiche, S.B. Hedges, M. Cummings, J. Wilkinson, H. Akashi, A. Filipinski, and an anonymous reviewer for comments on earlier drafts. This work is supported in part by research grants from the National Science Foundation to J.J.E., W.F.F., and S.K. and from the National Institutes of Health to S.K. There are no competing financial interests.

Literature Cited

- Aerts R, Chapin FS. 2000. The mineral nutrition of wild plants revisited: a re-evaluation of processes and patterns. *Adv Ecol Res* 30:1–67.
- Akashi H. 2003. Translational selection and yeast proteome evolution. *Genetics* 164:1291–303.
- Akashi H, Gojobori T. 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci USA* 99:3695–700.

- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–402.
- Baudouin-Cornu P, Schuerer K, Marlie P, Thomas D. 2004. Intimate evolution of proteins: proteome atomic content correlates with genome base composition. *J Biol Chem* 279:5421–8.
- Baudouin-Cornu P, Surdin-Kerjan Y, Marliere P, Thomas D. 2001. Molecular evolution of protein atomic composition. *Science* 293:297–300.
- Bragg JG, Hyder CL. 2004. Nitrogen versus carbon use in prokaryotic genomes and proteomes. *Proc R Soc Lond B* 271: S374–7.
- Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci USA* 96:4482–7.
- Elser JJ, Fagan WF, Denno RF, et al. (12 co-authors). 2000. Nutritional constraints in terrestrial and freshwater food webs. *Nature* 408:578–80.
- Hilborn R, Mangel M. 1997. *The ecological detective: confronting models with data*. Princeton, NJ: Princeton University Press.
- Lehninger AL, Nelson DL, Cox MM. 1993. *Principles of biochemistry*. New York: Worth Publishers.
- Lercher MJ, Urrutia AO, Hurst LD. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet* 31:180–3.
- Sokal RR, Rohlf FJ. 1994. *Biometry: the principles and practice of statistics in biological research*. New York, NY: W.H. Freeman & Company.
- Sonnhammer EL, Koonin EV. 2002. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet* 18:619–20.
- Sterner RW, Elser JJ. 2002. *Ecological stoichiometry: the biology of elements from molecules to the biosphere*. Princeton, NJ: Princeton University Press.
- Subramanian S, Kumar S. 2004. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics* 168:373–81.
- Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–80.
- Waterston RH, Lindblad-Toh K, Birney E, et al. (222 co-authors). 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–62.
- White TCR. 1993. *The inadequate environment: nitrogen and the abundance of animals*. New York: Springer-Verlag.

Manolo Gouy, Associate Editor

Accepted July 12, 2006